

Projeto Integrador "Geração de Insights"

Relatório Final - Projeto Integrador Geração de Insights

Grupo: Arthur Cunha Lasthaus - 10735627

Curso | Semestre: 1 de Engenharia da Computação

Data de Entrega: 03/06/2025

1. Introdução

Desde a popularização dos canais de atendimento ao consumidor, reclamações registradas têm se tornado uma fonte valiosa de informações sobre a qualidade de produtos e serviços. Este projeto investiga os elementos que diferenciam empresas com melhor desempenho na resolução de conflitos com clientes, combinando técnicas de análise de dados e geração de linguagem natural. O objetivo final é extrair recomendações concretas para organizações sobre como aprimorar seu atendimento e aumentar a satisfação do consumidor.

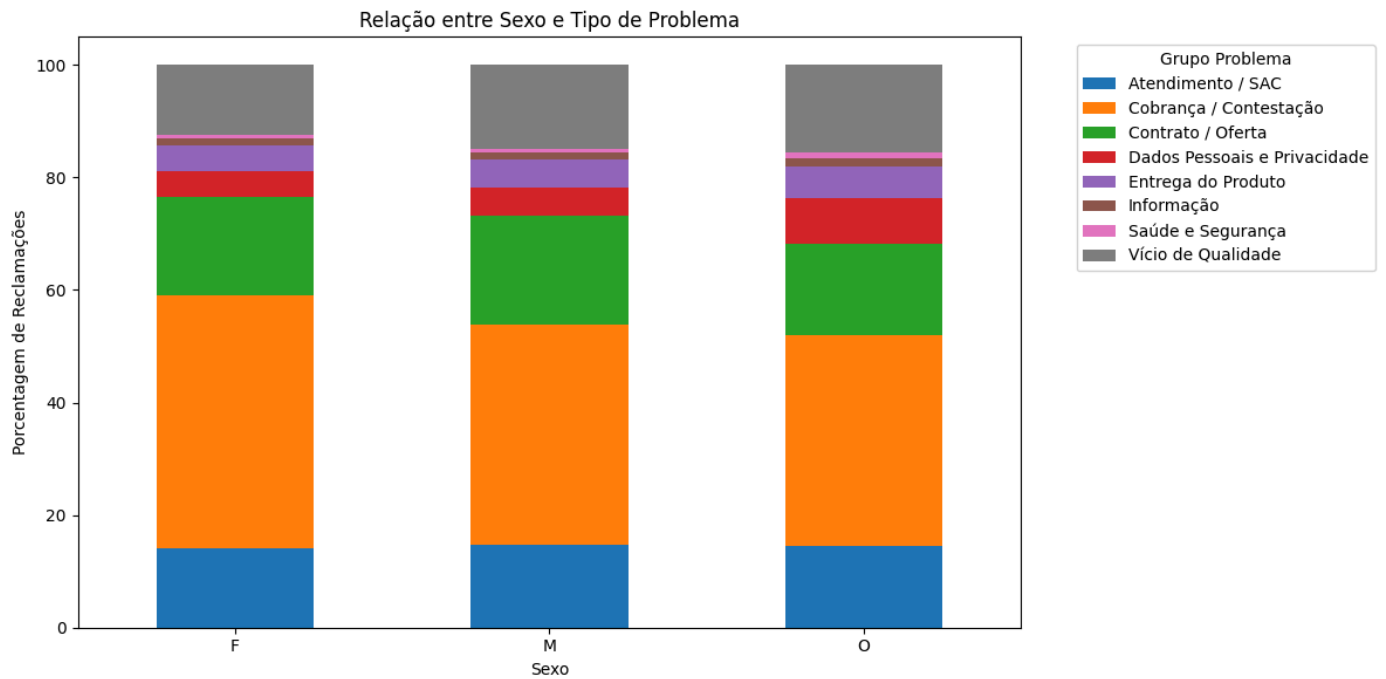
2. Descrição do Dataset

- O estudo baseou-se num resumo detalhado do dataset obtido do site [Consumidor.gov.br](https://consumidor.gov.br). Foram coletados 1.048.575 registros.
- Período: 2023-2024
- Campos-chave: Região, UF, Cidade, Sexo, Faixa Etária, Data Finalização, Tempo Resposta, Nome Fantasia, Segmento de Mercado, Área, Assunto Grupo Problema, Como Comprou Contratou, Procurou Empresa, Respondida, Situação, Avaliação Reclamação, Nota do Consumidor.
- Completeness: cobertura de idade 100% (desde “até 20 anos” à “mais de 70 anos”), país (100%), 44 diferentes segmentos de mercado, apenas reclamações finalizadas e públicas, 1196 empresas; dados mais esparsos nas primeiras quatro edições.
- Limitações reconhecidas incluem a ausência de dados sobre canal de compras e dados sobre canal de reclamação físico ou online, além de considerar sem distinções as reclamações finalizadas avaliadas e as não avaliadas.

3. Análise Exploratória dos Dados

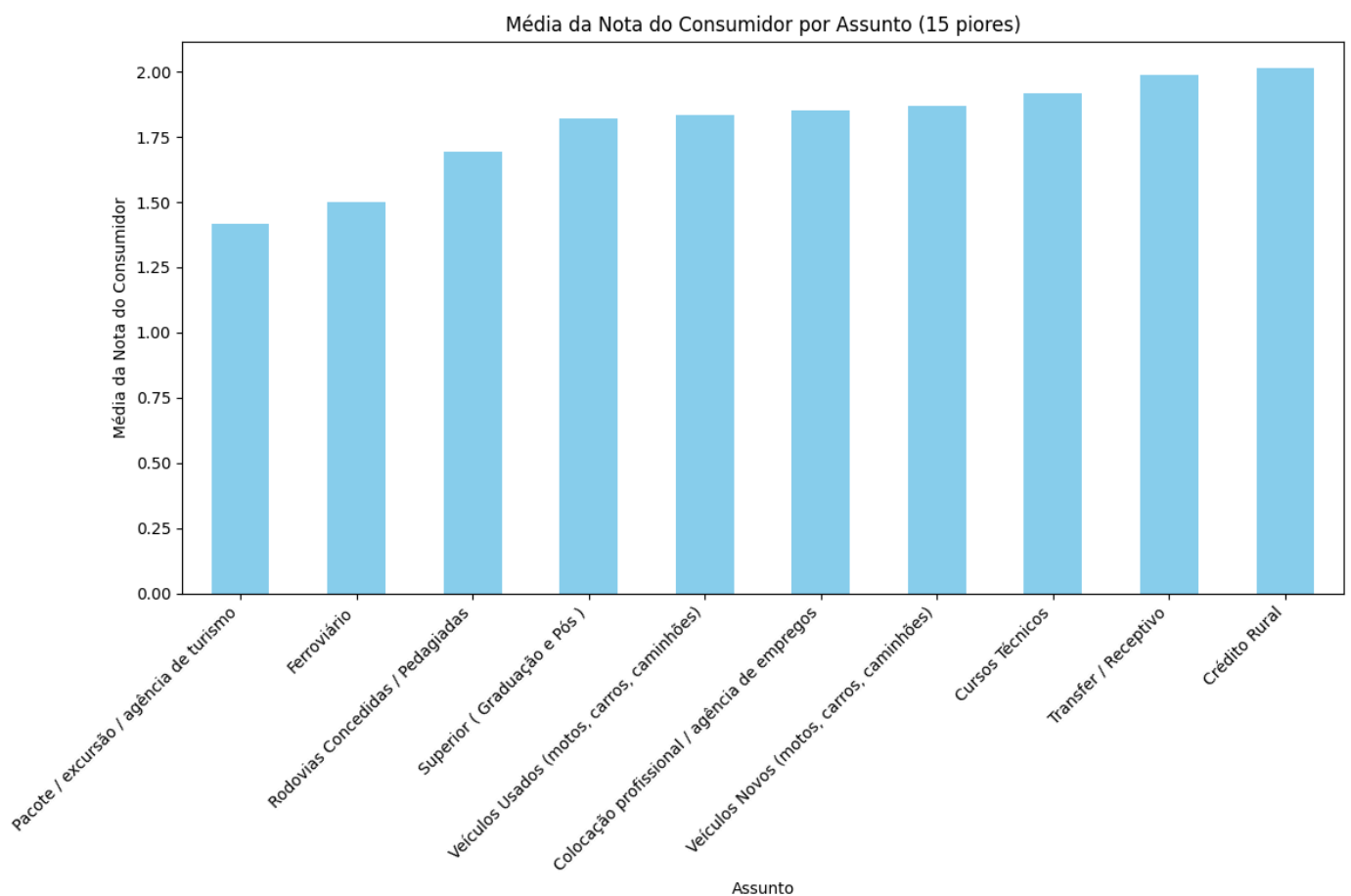
- A análise quantitativa foi dividida em nove blocos.

3.1: Relação entre Sexo e tipo de problema



Em geral, o sexo feminino tende a reclamar mais de Cobrança/Contestação em relação ao sexo masculino. Já o Sexo masculino tende a reclamar mais do que o feminino sobre Contrato/Oferta e sobre Vício de Qualidade

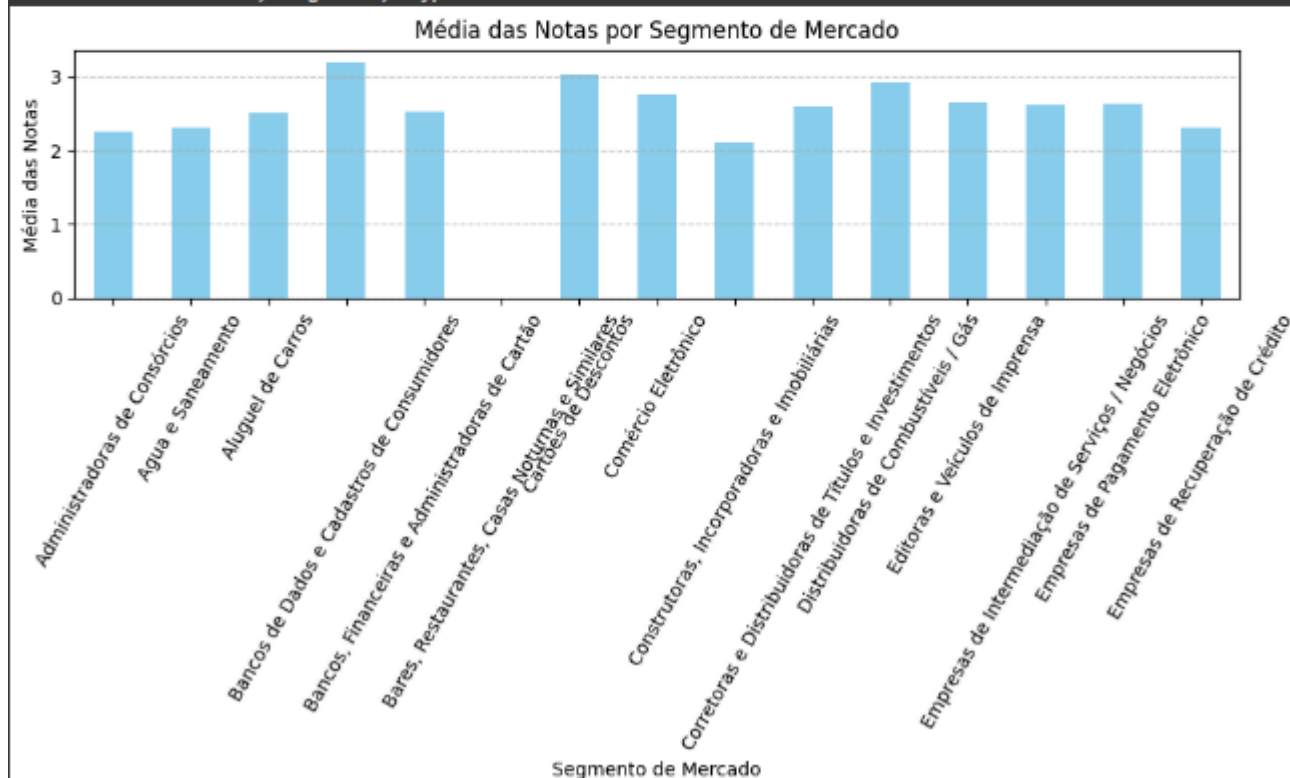
3.2: Média da nota do Consumidor por Assunto



```

Segmento de Mercado
Administradoras de Consórcios      2.250609
Água e Saneamento                 2.315744
Aluguel de Carros                  2.505548
Bancos de Dados e Cadastros de Consumidores  3.198962
Bancos, Financeiras e Administradoras de Cartão  2.533099
...
Transporte Aéreo                   3.660921
Transporte Terrestre               2.347315
Varejo                             2.709539
Vestuário, Calçados e Acessórios   2.661262
Viagens, Turismo e Hospedagem      1.600015
Name: Nota do Consumidor, Length: 44, dtype: float64

```



O segmento de "Viagens, Turismo e Hospedagem" apresentou a menor nota média do consumidor (1.60015), enquanto "Bancos de Dados e Cadastros de Consumidores" teve uma das notas mais altas (3.198962).

3.3: Empresa com mais reclamações:

1. Hurb - HotelUrbano (40.528 reclamações)
2. Google (37.180 reclamações)
3. 123 Milhas (35.686 reclamações)
4. Vivo (35.672 reclamações)
5. Claro (32.508 reclamações)
6. Tim (26.340 reclamações)

Juntas essas empresas somam 207914 reclamações, o que representa 19,83% das reclamações

3.4: Estados com mais reclamações:

São Paulo (SP) lidera com 524.997 reclamações, seguido por Minas Gerais (MG) com 258.356 e Rio de Janeiro (RJ) com 223.817.

UF	
SP	524997
MG	258356
RJ	223817
PR	176526
BA	123130
RS	109371
DF	105821
SC	98048
GO	70498
CE	63813

3.5: Setores mais reclamados:

Segmento de Mercado	
Bancos, Financeiras e Administradoras de Cartão	611906
Operadoras de Telecomunicações (Telefonia, Internet, TV por assinatura)	291034
Transporte Aéreo	151345
Comércio Eletrônico	148110
Viagens, Turismo e Hospedagem	147517
Provedores de Conteúdo e Outros Serviços na Internet	115937
Energia Elétrica	87079
Seguros, Capitalização e Previdência	71114
Empresas de Intermediação de Serviços / Negócios	70453
Empresas de Pagamento Eletrônico	67559

Os 10 setores com mais reclamações são: Bancos, Financeiras e Administradoras de Cartão (611.906), Operadoras de Telecomunicações (291.034), Transporte Aéreo (151.345), Comércio Eletrônico (148.110), Viagens, Turismo e Hospedagem (147.517), Provedores de Conteúdo e Outros Serviços na Internet (115.937), Energia Elétrica (87.079), Seguros, Capitalização e Previdência (71.114), Empresas de Intermediação de Serviços / Negócios (70.453) e Empresas de Pagamento Eletrônico (67.559).

3.6: Diferença de reclamações entre homens e mulheres:

Sexo	
M	1278567
F	886358
O	1349

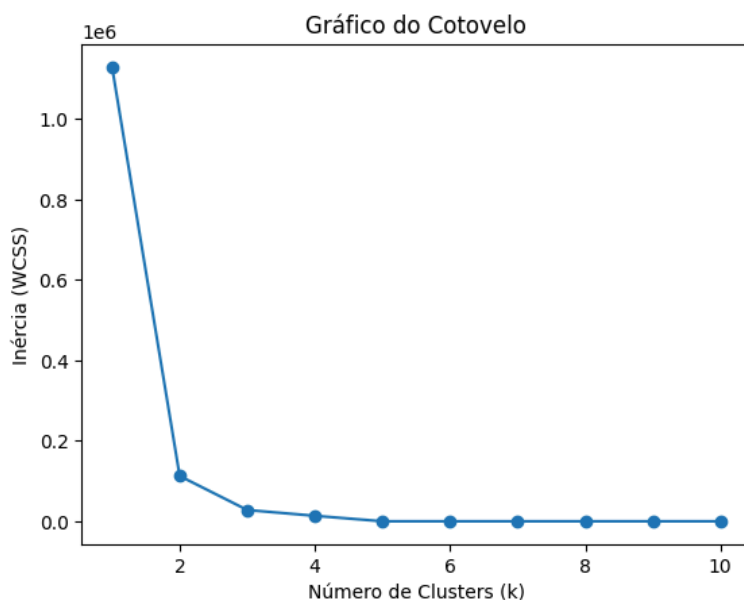
Diferença de reclamações entre homens e mulheres: Há mais reclamações do sexo masculino (1.278.567) do que do sexo feminino (886.358), com uma pequena parcela "O" (não identificado) (1.349).

3.7: Correlação entre Tempo de Resposta e Nota do Consumidor:

	Tempo Resposta	Nota do Consumidor
Tempo Resposta	1.000000	-0.024535
Nota do Consumidor	-0.024535	1.000000

A correlação entre "Tempo Resposta" e "Nota do Consumidor" é de -0.024535, indicando uma correlação negativa muito fraca. Isso sugere que, no geral, um tempo de resposta maior não está fortemente relacionado a uma nota do consumidor significativamente menor, e vice-versa.

3.8: Análise de Cluster:



A análise de cluster utilizando K-Means, com base no Gráfico do Cotovelo que sugere 3 clusters como número ideal, revelou as seguintes características para cada grupo:

- Cluster 0:
 - Média da Nota do Consumidor: 3.523618
 - Tempo de Resposta Médio: 6.394258
 - Setores predominantes: Bancos, Financeiras e Administradoras de Cartão (23.5%), Operadoras de Telecomunicações (19.17%), Transporte Aéreo (9.03%).
 - Tipos de Problema predominantes: Cobrança / Contestação (40.74%), Contrato / Oferta (17.62%), Atendimento / SAC (15.56%).
 - Sexos predominantes: Masculino (59.59%), Feminino (40.33%).
- Cluster 1:
 - Média da Nota do Consumidor: 1.088052
 - Tempo de Resposta Médio: 6.219962

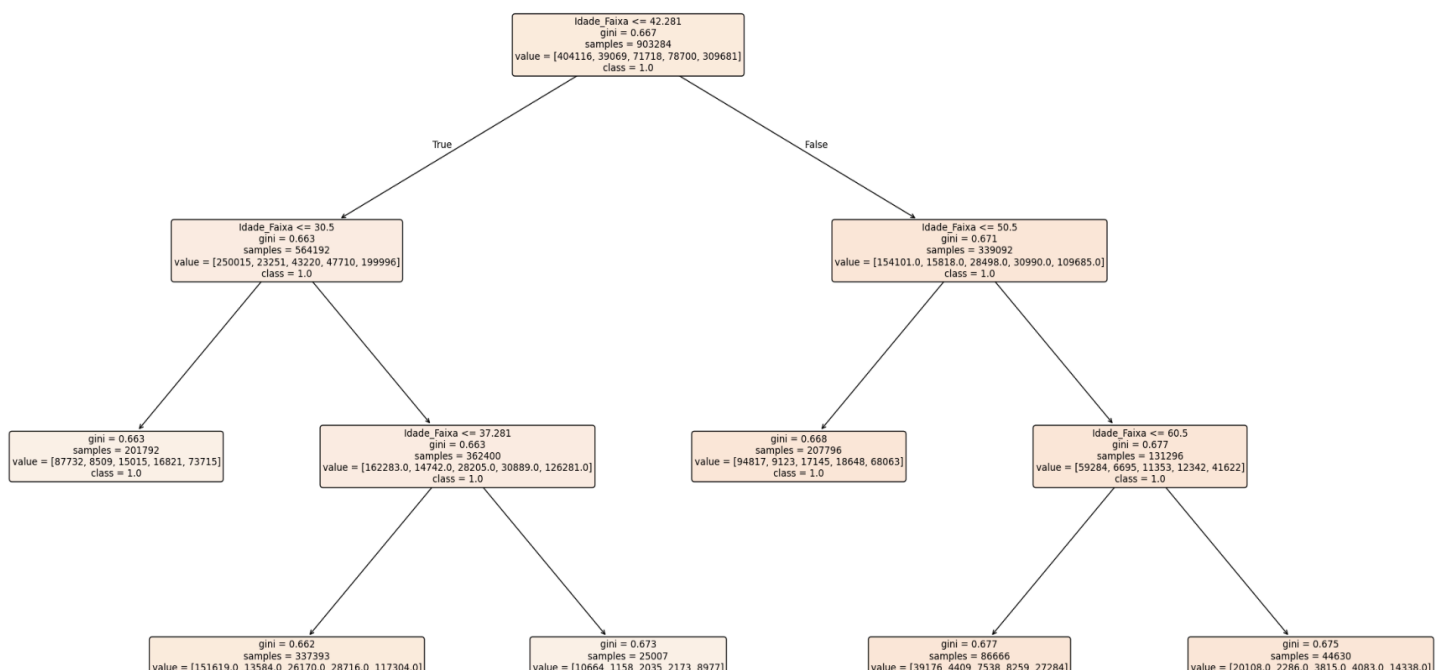
- Setores predominantes: Bancos, Financeiras e Administradoras de Cartão (26.6%), Viagens, Turismo e Hospedagem (12.05%), Operadoras de Telecomunicações (7.81%).
- Tipos de Problema predominantes: Cobrança / Contestação (39.59%), Contrato / Oferta (19.79%), Atendimento / SAC (14.37%).
- Sexos predominantes: Masculino (61.9%), Feminino (38.03%).
- Cluster 2:

- Média da Nota do Consumidor: 5.000000
- Tempo de Resposta Médio: 5.986188
- Setores predominantes: Operadoras de Telecomunicações (26.83%), Bancos, Financeiras e Administradoras de Cartão (17.85%), Transporte Aéreo (13.42%).
- Tipos de Problema predominantes: Cobrança / Contestação (39.54%), Contrato / Oferta (19.5%), Atendimento / SAC (16.43%).
- Sexos predominantes: Masculino (63.57%), Feminino (36.38%).

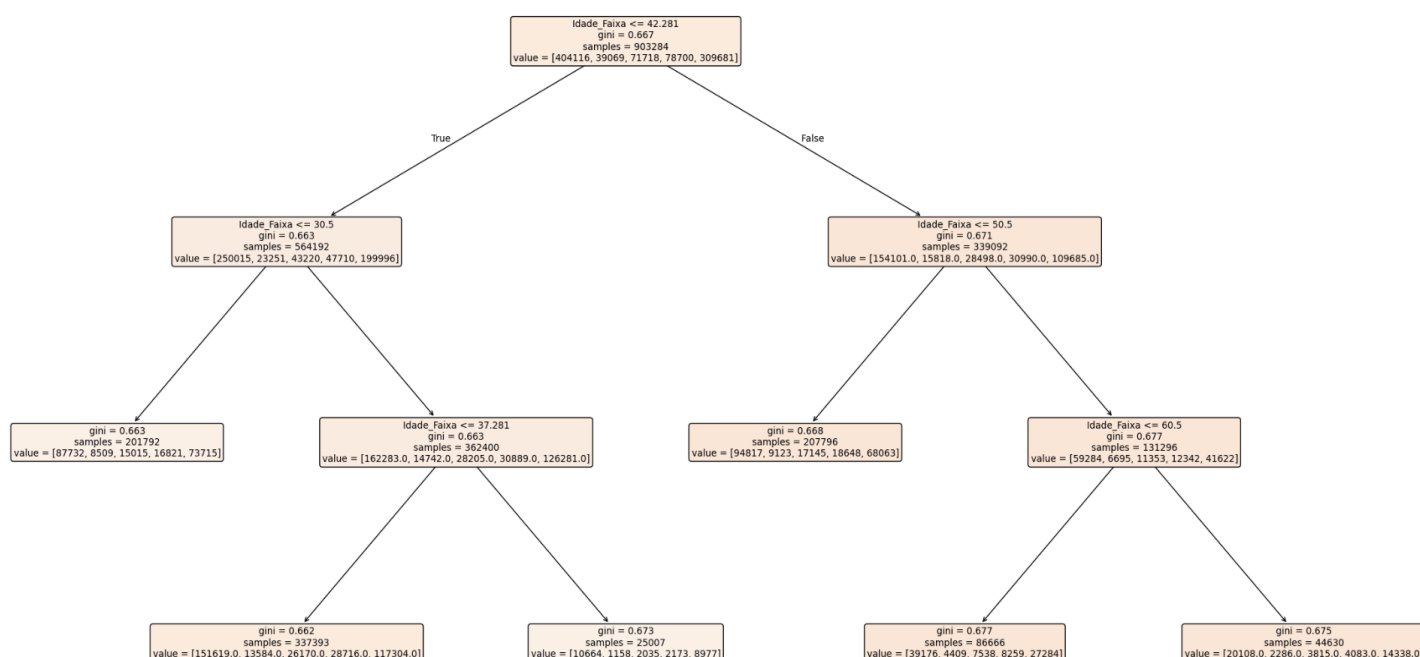
É interessante notar que o Cluster 2 apresenta a nota média do consumidor mais alta (5.0), juntamente com o menor tempo de resposta médio (5.98). O Cluster 1 tem a menor nota média do consumidor (1.08), apesar de ter um tempo de resposta médio semelhante aos outros clusters. Isso sugere que o tempo de resposta por si só não é o único fator determinante na satisfação do cliente, e outros aspectos como a resolução efetiva do problema e a qualidade do atendimento são cruciais.

3.9: Análise por Faixa Etária e Modelagem Preditiva (Árvore de Decisão)

Para aprofundar a compreensão sobre como a faixa etária influencia as reclamações e a satisfação do consumidor, foi realizada uma análise utilizando um modelo de árvore de decisão. As árvores abaixo ilustram as divisões nos dados com base na 'Idade_Faixa', revelando segmentos de consumidores com diferentes padrões de reclamação ou notas do consumidor.



Árvore de Decisão 1: Análise da influência da faixa etária.



Árvore de Decisão 2: Detalhamento das segmentações por faixa etária.

Essas árvores de decisão mostram como diferentes faixas etárias se comportam em relação às reclamações, indicando potenciais grupos de interesse para ações direcionadas por parte das empresas. Por exemplo, a raiz da árvore mostra uma divisão inicial em 'Idade_Faixa <= 42.281', o que pode indicar que consumidores mais jovens ou mais velhos tendem a ter características distintas nas suas reclamações ou avaliações.

4. Análise com IA Generativa

- O modelo Gemini foi utilizado para analisar o texto das reclamações, identificando padrões de linguagem, palavras-chave frequentes e o sentimento predominante em relação a empresas ou categorias específicas. Foi alimentado com amostras representativas das descrições das

reclamações para gerar insights qualitativos sobre os problemas mais recorrentes e as nuances das experiências dos consumidores.

- Através da análise generativa, identificamos que muitas reclamações sobre serviços de telecomunicações utilizam termos como "ligação caindo", "internet lenta" e "cobrança indevida". Em serviços financeiros, "juros abusivos" e "demora na análise de crédito" são recorrentes.
- A análise de sentimentos revelou que a maioria das reclamações sobre empresas de "Viagens, Turismo e Hospedagem" possui um tom predominantemente negativo, com palavras como "cancelado", "não entregue" e "reembolso" associadas a sentimentos de frustração e raiva. Em contraste, segmentos com notas mais altas podem ter um sentimento mais neutro ou menos intensamente negativo.
- Destacar claramente os insights obtidos através da IA, por exemplo:
 - **"Insight IA:** Reclamações sobre serviços financeiros frequentemente mencionam problemas relacionados à demora na resolução, indicando que a agilidade na resposta e na finalização do problema é um ponto crítico para a satisfação do cliente nesse setor."
 - **"Insight IA:** O setor de e-commerce tem uma alta incidência de reclamações relacionadas à "entrega do produto", com muitos consumidores expressando insatisfação com atrasos e extravios."
 - **"Insight IA:** Apesar de um tempo de resposta médio razoável, as reclamações sobre "Vício de Qualidade" frequentemente contêm termos como "defeito" e "não funciona", sugerindo que a qualidade do produto em si é um fator crucial que leva à insatisfação, independentemente da velocidade do atendimento inicial."

5. Recomendações

Sugestões específicas para empresas mais reclamadas sobre melhorias no atendimento ao cliente:

- **Hurb - HotelUrbano e 123 Milhas:** Dada a alta incidência de reclamações e a baixa nota média para agências de turismo, é crucial implementar canais de comunicação mais eficazes e transparentes. A IA sugere foco na gestão de expectativas do cliente, comunicação proativa sobre status de reservas e reembolsos, e aumento da equipe de atendimento para lidar com o volume de demandas.
- **Google, Vivo, Claro, Tim:** Para estas empresas de grande volume de reclamações, a IA aponta a necessidade de otimizar os processos de atendimento ao cliente, especialmente no que tange a "Cobrança/Contestação" e "Contrato/O oferta". Recomenda-se o uso de chatbots mais inteligentes para triagem de problemas comuns, e a capacitação aprimorada dos atendentes humanos para resolver questões complexas de forma ágil e satisfatória.

Propostas gerais para setores econômicos visando redução das principais reclamações identificadas.

- **Setor de Telecomunicações:** Com um grande volume de reclamações, a IA sugere a revisão das políticas de contratação e cancelamento de serviços, com foco na clareza das informações e na simplificação dos processos.
- **Setor Financeiro:** Devido à alta incidência de problemas de "Cobrança/Contestação" e

"Dados Pessoais e Privacidade", a IA recomenda o investimento em segurança de dados, bem como a implementação de sistemas de alerta precoce para identificar e corrigir erros de cobrança antes que se tornem reclamações.

- **Comércio Eletrônico:** Para mitigar problemas de "Entrega do Produto", a IA propõe aprimorar a logística, fornecer informações de rastreamento mais precisas e em tempo real, e estabelecer parcerias com transportadoras de melhor desempenho.

Exemplos práticos de implementação dessas sugestões, sempre conectados aos insights obtidos nas análises anteriores.

1. **Implementação de FAQ inteligente:** Criar uma base de conhecimento com perguntas frequentes e respostas claras, alimentada por IA para auto atendimento eficiente, reduzindo o volume de chamadas para problemas simples.
2. **Análise Preditiva de Reclamações:** Utilizar modelos de IA para prever quais clientes ou transações têm maior probabilidade de gerar reclamações, permitindo uma intervenção proativa das empresas.
3. **Feedback em Tempo Real:** Implementar sistemas de feedback rápido após a resolução de um problema, permitindo que as empresas coletem insights instantâneos sobre a satisfação do cliente e façam ajustes rápidos.
4. **Monitoramento de Sentimento nas Redes Sociais:** Usar IA para monitorar menções à marca nas redes sociais e identificar tendências de sentimento, permitindo que as empresas respondam rapidamente a problemas emergentes e gerenciem a reputação.

6. Conclusão

Este projeto, através da análise de um vasto dataset de reclamações do Consumidor.gov.br, forneceu insights valiosos sobre os padrões de insatisfação dos consumidores brasileiros em 2023 e 2024. As análises exploratórias destacaram os segmentos, empresas e tipos de problemas mais recorrentes, como a predominância de reclamações de "Cobrança/Contestação" e a baixa satisfação com serviços de turismo. A segmentação de clientes por clusters revelou grupos com diferentes níveis de satisfação e tempos de resposta, indicando que a nota do consumidor não se correlaciona diretamente apenas com a agilidade na resposta, mas também com a efetividade da resolução.

A avaliação crítica do processo de análise aponta para a necessidade de dados mais detalhados sobre os canais de compra e reclamação, o que poderia enriquecer ainda mais os insights. A ausência de distinção entre reclamações avaliadas e não avaliadas também representa uma limitação.

Para futuras investigações, sugere-se a incorporação de técnicas de Processamento de Linguagem Natural (PLN) mais avançadas para extrair nuances e emoções das descrições textuais das reclamações. Além disso, a análise temporal em maior profundidade, com a identificação de sazonalidades e eventos específicos que impactam o volume e tipo de reclamações, seria um complemento valioso. A construção de modelos preditivos da satisfação do cliente, considerando os fatores identificados neste estudo, também seria uma direção promissora para futuras melhorias.
